智慧科技與應用統計學報
第 9 卷 第 2 期
2011 年 12 月
31-48 頁

異質性多變量 *t* 母體分類法之研究

陳屏儒 [1] 郭于銓 [1] 林宗儀 [1, 2*]

[1] 國立中興大學應用數學系
[2] 國立中興大學統計學研究所

摘 要

在多變量 *t* 分佈的假設下，我們針對多群異質母體建立新的分類法則。這種推廣相較於被經常使用的多變量常態假設會有較穩健的性質。在新的分類架構下，未知參數的估計值是透過期望最大化演算法(EM algorithm)去求取。相較於常態分佈，當收集到的資料無法避免地具有厚尾或離群值時，所提出的這項分類技術會特別有用。實驗結果顯示，新的分類規則在某些情況下會比傳統的區別分析法則有更好的表現，尤其是當資料的分佈遠離常態。

關鍵詞：區別分析；EM 演算法；最大概似估計；多變量常態分佈；多變量 *t* 分佈

*通訊作者聯絡方式：tilin@amath.nchu.edu.tw

Classification Rules for Heterogeneous Multivariate *t* Populations

Ping-Ju Chen[1], Yu-Chuan Kuo[1], and Tsung-I Lin[1, 2]

[1] Department of Applied Mathematics, National Chung Hsing University, Taiwan
[2] Institute of Statistics, National Chung Hsing University, Taiwan

Abstract

We establish novel classification rules for heterogeneous populations under the assumption of multivariate *t* distribution, which can be treated as a robust generalization of the routinely used multivariate normal one. In this new classification framework, the unknown parameters are estimated by the maximum likelihood method via the expectation maximization (EM) algorithm. The proposed classification technique is particularly useful when the collected data unavoidably contain longer than normal tails or outlying observations. Experimental results show that the new classification rules may outperform the traditional classifiers in some scenarios, especially when the underlying distribution of data is far from normal.

Key words: discriminant analysis; EM algorithm; maximum likelihood estimation; multivariate normal distribution; multivariate *t* distribution

1. 緒論

區別分析(discriminant analysis)是多變量統計分析中用於判別樣本資料所屬其類型或族群的一種方法，與集群分析相同的是將相似的樣本資料歸爲一類族群，不同處卻在於集群分析預先不知道研究對象之所屬類別，而區別分析是在研究對象所屬類別已知的情況下，根據樣本資料推導出一個或一組區別函數，同時指定一種分類規則，用於確定待判別樣本的所屬類別。

在古典的區別分析中，由於統計模型推論上的便利與大樣本性質，常常對於資料的母體假設爲常態分佈，但當資料具有離群值時，這些離群值會對母體假設爲常態分佈下所架構的區別函數造成極大的影響。有鑒於此，在兩群資料爲多變量 *t* 分佈且具相同共變異數與相同自由度假設下，Sutradhr (1990)利用樣本統計量去推估 *t* 分佈母體的參數，建構一個新的區別函數。有關多變量 *t* 分佈的應用與推論，可參閱 Kotz and Nadarajah (2004)。

對於多群異質性母體，本文探討如何處理具離群值的資料並架構比常態分佈假設下更穩健的區別函數。在母體爲多變量 *t* 分佈假設下，本文推廣 Sutradhr (1990)的方法，同時考慮四種不同的參數結構，利用期望最大化(EM)演算法(Dempster et al., 1977)去計算參數的最大概似估計值，建構多群 *t* 分佈母體的區別法則。

本篇論文架構如下，第二節爲文獻探討，分別回顧區別分析與多變量 *t* 分佈。第三節爲研究方法，在母體是多變量常態分佈與多變量 *t* 分佈的假設下，我們建立 6 種分類法則，在多變量常態分佈假設下，考慮共變異數矩陣相等或不相等的兩種參數結構；在多變量 *t* 分佈假設下，則分別考慮共變異數矩陣相等或不相等與自由度相等或不相等的四種參數結構。第四節，透過模擬與實例分析比較 6 種分類法則的表現。第五節爲結論。

2. 文獻探討

2.1 區別分析(discriminant analysis)

區別分析又稱爲鑑別分析或判別分析(與集群分析不同)，其目的在了解群體之間的差異，並將預測的觀察值做分類，辨別觀測值所屬的群組。區別分析利用區別變數建立一個區別函數值或區別規則做爲分類的準則。新的觀測值依照其區別函數值或區別規則做分類，使新的觀測值之分類與原本的資料類別最爲接近。

在多變量分析中，區別分析爲判別資料類別的一種統計方法，其利用數個區別變數去預測資料的類別。運用區別分析，可以決定較適合的區別變數，並依照其區別規則對數個異質群體做分類，讓預測的資料分類後較接近原始資料的類別，使分錯率最小。在考慮母體服從多變量 *t* 分佈下，其條件有共變異數相同與否、以及自由度相同與否。

而區別規則在區別分析中是很重要的一環，有分類依據的區別規則，才能對新的觀測值做群體歸類。建立區別規則可以根據新的觀測值會發生在某個群組的後驗機率，然後將新的觀測值區分在發生後驗機率最大的群組。

2.2 多變量 *t* 分佈(multivariate *t* distribution)

假設 Y 是一個 p 維度的隨機向量並且服從多變量 t 分佈，表示爲 $t_p(\mu, \Sigma, v)$ ，當中 μ 爲平均數向量，Σ 爲共變異數矩陣，v 爲自由度。根據 Kotz and Nadarajah (2004)，隨機向量 Y 可以表示如下：

$$Y = \mu + \frac{X}{\sqrt{\tau}}, \quad X \sim N_p(0, \Sigma), \quad \tau \sim Gamma\left(\frac{v}{2}, \frac{v}{2}\right), \quad X \perp \tau .$$

Y 的邊際分佈爲

$$f(Y) = \frac{1}{|\Sigma|^{\frac{1}{2}}(\pi v)^{\frac{p}{2}}\Gamma(\frac{v}{2})}\Gamma\left(\frac{v+p}{2}\right)\left(1+\frac{(Y-\mu)'\Sigma^{-1}(Y-\mu)}{v}\right)^{-\frac{v+p}{2}} .$$

利用貝氏定理，可以得到的 $\tau|Y$ 後驗分佈爲

$$\tau|Y \sim Gamma\left(\frac{v+p}{2}, \frac{(Y-\mu)'\Sigma^{-1}(Y-\mu)+v}{2}\right).$$

在統計分析中，因多變量 t 分佈的尾部較豐厚，能夠涵蓋離群值的資料。厚尾的多變量 t 分佈提供了一種在配適資料上穩健的特性。其中自由度 v 是控制分佈尾巴的厚度。而當 v 越來越小時，分佈尾巴會越來越厚，分佈覆蓋的空間就越大，能有效防止離群點對各推論值的影響。

以下藉由多變量常態分佈與多變量 t 分佈所畫的二維度聯合等高線圖來了解兩者的差異。其聯合等高線圖的機率密度函數爲 $f(x; \mu, \Sigma, v) = c$ ，平均數向量爲 $\mu = 0$ ，共變異數矩陣爲 $\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$ 以及三個不同的自由度分別爲 $v = 3$ ， $v = 20$ 和 $v = \infty$ 。其中 $v = \infty$ 相當於多變量常態分佈。我們考慮 $c = 0.1$ 和 $c = 0.01$ 兩種情況，如圖 1。

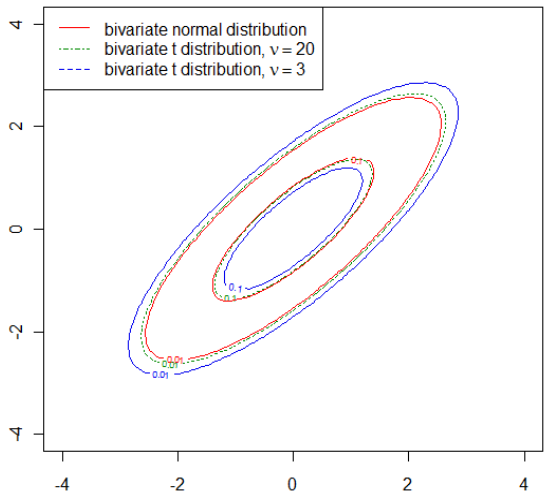

圖 1. 聯合等高線圖(機率密度等於 0.1 和 0.01)

在圖 1 中，實線表多變量常態分佈的機率密度的等高線；短虛線與點線分別爲多變量 t 分佈的機率密度函數的等高線，其中自由度 v 分別爲 3 與 20。當機率密度函數爲 0.1 時，可以明顯看出多變量常態分佈所涵蓋的資料較多變量 t 分佈更爲廣泛。相反地，當機率密度函數爲 0.01 時，可以發現多變量 t 分佈所涵蓋的資料較多變量常態分佈更爲廣泛。

此外，當自由度爲 $v=20$ 的點線很靠近實線，表示當多變量 t 分佈的自由度越大時，越接近多變量常態分佈。總而言之，當多變量 t 分佈的自由度越小且機率密度函數值越小時，所涵蓋的資料範圍越廣。

3. 研究方法

給定 $f_i(x)$ 爲母體 P_i 的機率密度且假設 π_i 爲母體 P_i 的事前機率，$i=1,\ldots,g$ ；$c(k|i)$ 爲觀測值 x 來自 P_i ，但被分錯到 P_k 所需的成本，$k,i=1,\ldots,g$ 。對於 $k=i$ ，$c(i|i)=0$ ，給定 Ω_k 是 x 分類到 P_k 的集合，及

$$P(k|i)=P(x\in\Omega_k|P_i)=\int_{\Omega_k} f_i(x)dx,\quad k,i=1,\ldots,g,$$
$$P(i|i)=1-\sum_{\substack{k=1\\k\neq i}}^{g}P(k|i).$$

對於分錯一個 x 從 P_1 到 P_2 或 $P_3,\ldots,$或 P_g ，這條件的預期成本爲

$$\begin{aligned}ECM(1)&=P(2|1)c(2|1)+P(3|1)c(3|1)+\cdots+P(g|1)c(g|1)\\&=\sum_{k=2}^{g}P(k|1)c(k|1).\end{aligned}$$

經由上述，可以獲得條件的分錯預期成本 $ECM(2),\ldots,ECM(g)$ ，再將事前機率乘以每一個條件的 ECM ，並且加總，得到 ECM 爲：

$$\begin{aligned}ECM&=\pi_1 ECM(1)+\pi_2 ECM(2)+\cdots+\pi_g ECM(g)\\&=\sum_{i=1}^{g}\pi_i\left(\sum_{\substack{k=1\\k\neq i}}^{g}P(k|i)c(k|i)\right).\end{aligned}$$

其中，母體 P_i 的事前機率爲 π_i 。

我們可以選擇一個互斥的分類區域 $\Omega_1,\ldots,\Omega_g$ ，找到最理想的分類法則，使得 ECM 最小化。這分類區域，爲分配觀測值 x 到母體 P_k ，$k=1,\ldots,g$ ，使 ECM 最小化，其被定義爲

$$\sum_{\substack{i=1\\i\neq k}}^{g}\pi_i f_i(x)c(k|i),$$

爲最小。假設所有分錯的成本都相等，在這種情況下，ECM 最小化規則是分錯的總機率最小化 (minimum total probability of misclassification, TPM)規則。沒有一般性的損失，可以定義所有分錯成本等於 1。當分配 x 到母體 P_k ，$k=1,\ldots,g$ ，可以得到

$$\sum_{\substack{i=1 \\ i\neq k}}^{g} \pi_i f_i(x),$$

爲最小。

在數個母體下，根據最佳分類法則定理(Johnson and Wichern, 2007)，當分錯成本是一樣時，可以得到 *ECM* 最小化分類法則爲：

分配觀測值 x_0 到 P_k ，假使

$$\ln(\pi_k f_k(x_0)) > \ln(\pi_i f_i(x_0)), \text{ 所有 } i \neq k. \tag{1}$$

3.1 異質性多變量常態母體假設之分類法則

定義 $f_i(x)$ 爲多變量常態隨機向量 x 其平均數向量爲 μ_i 和共變異數矩陣爲 Σ_i 之密度函數，爲

$$f_i(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\right], \ i=1,\ldots,g.$$

假使 $c(i|i)=0$ 與 $c(k|i)=1$ ， $k \neq i$ 。則(1)式中的分類法則爲：

分配 x 到 P_k ，假使

$$\begin{aligned} \ln(\pi_k f_k(x)) &= \ln \pi_k - \left(\frac{p}{2}\right)\ln(2\pi) - \frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x-\mu_k)'\Sigma_k^{-1}(x-\mu_k) \\ &= \max \ln(\pi_i f_i(x)). \end{aligned} \tag{2}$$

在(2)式中，常數 $(p/2)\ln(2\pi)$ 可以被忽略不計。因此，定義二次式的區別分數(quadratic discrimination score)於第 i 個母體爲

$$d_i^Q(x) = -\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i) + \ln \pi_i, \ i=1,\ldots,g. \tag{3}$$

參數結構 1：當共變異數矩陣不相等。

在區別分析中，可以建構在常態母體其參數結構爲共變異數矩陣不相等之下，TPM 最小化的分類法則爲：

分配 x 到 P_k ，假使

$$\text{二次式的分數 } d_k^Q(x) = \max\left\{d_1^Q(x), d_2^Q(x),\ldots,d_q^Q(x)\right\},$$

其中 $d_i^Q(x)$ 被給定於(3)式。

實際上，平均數向量 μ_i 和共變異數矩陣 Σ_i 都是未知的，所以分類法必須修改。我們用樣本資料來估計母體參數，這些資料的樣本平均數向量和共變異數矩陣爲：

$\bar{x}_i$ = 樣本平均數向量， S_i = 樣本共變數矩陣， n_i = 樣本大小。

常以樣本統計量 $\bar{x}_i$ 和 S_i 代替(3)式中母體的 μ_i 和 Σ_i ，則二次式的區別分數 $d_i^Q(x)$ 的估計量 $\hat{d}_i^Q(x)$ 爲

$$\hat{d}_i^Q(x) = -\frac{1}{2}\ln|S_i| - \frac{1}{2}(x-\bar{x}_i)'S_i^{-1}(x-\bar{x}_i) + \ln\pi_i, \quad i=1,\ldots,g. \tag{4}$$

而分類法則以樣本爲基礎，建構在數個常態母體其參數結構爲共變異數矩陣不相等之下，估計 TPM 最小化的分類法則爲：

分配 x 到 P_k ，假使

$$\text{二次式的分數 } \hat{d}_k^Q(x) = \max\left\{\hat{d}_1^Q(x), \hat{d}_2^Q(x),\ldots,\hat{d}_q^Q(x)\right\}.$$

其中 $\hat{d}_i^Q(x)$ 被給定於(4)式。

參數結構 2：共變異數矩陣相等($\Sigma_1 = \cdots = \Sigma_g = \Sigma$)。

假使母體的共變異數矩陣 Σ_i 是相等的，此區別分數在(3)式中會變成

$$d_i^Q(x) = -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}x'\Sigma^{-1}x + \mu_i'\Sigma^{-1}x - \frac{1}{2}\mu_i'\Sigma^{-1}\mu_i + \ln\pi_i.$$

對於 $d_1^Q(x), d_2^Q(x),\ldots,d_q^Q(x)$ ，前面兩項是一樣的，可以被省略。其次，定義線性區別分數 $d_i(x)$ 爲

$$d_i(x) = \mu_i'\Sigma^{-1}x - \frac{1}{2}\mu_i'\Sigma^{-1}\mu_i + \ln\pi_i, \quad i=1,\ldots,g.$$

這線性區別分數 $d_i(x)$ 的估計量 $\hat{d}_i(x)$ 是以 Σ 的混合估計量爲基礎。

$$S_{pooled} = \frac{(n_1-1)S_1 + (n_2-1)S_2 + \cdots + (n_g-1)S_g}{n_1+n_2+\cdots+n_g-g},$$

其中 $S_i = (1/n_i - 1)\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'$ ， $i=1,\ldots,g$ 。因此， $\hat{d}_i(x)$ 給定爲

$$\hat{d}_i(x) = \bar{x}_i'S_{pooled}^{-1}x - \frac{1}{2}\bar{x}_i'S_{pooled}^{-1}\bar{x}_i + \ln x_i, \quad i=1,\ldots,g. \tag{5}$$

其後，建構在常態母體其參數結構($\Sigma_1 = \cdots = \Sigma_g = \Sigma$)下，估計 TPM 最小化的分類法則爲：

分配 x 到 P_k ，假使

$$\text{線性區別分數 } \hat{d}_k(x) = \max\left\{\hat{d}_1(x), \hat{d}_2(x),\ldots,\hat{d}_g(x)\right\}$$

其中 $\hat{d}_i(x)$ 被給定於(5)式。

3.2 異質性多變量 *t* 母體假設之分類法則

在此節中探討更一般情況下多個母體爲多變量 t 分佈假設下最佳分類法則之建構。假設 $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ 爲來自第 i 個母體 $P_i \sim t_{p_i}(\mu_i, \Sigma_i, v_i)$ ， $i = 1, \ldots, g$ ，其中全部資料的樣本大小爲 $n = n_1 + n_2 + \cdots + n_g$ 。

透過多變量 *t* 分佈的性質，可以得到以下的階層表示式：

$$Y_{ij} \mid \tau_{ij} \sim N(\mu_i, \tau_{ij}^{-1}\Sigma_i), \quad \tau_{ij} \sim Gamma(\frac{v_i}{2}, \frac{v_i}{2}), \quad i = 1, \ldots, g, \quad j = 1, \ldots, n_i.$$

我們延伸 Liu (1997)所提出估計多變量 *t* 分佈參數的期望最大化 EM 演算法，推廣到多群 *t* 母體之不同參數結構下，建構 *ECM* 最小的最佳分類法則。

參數結構 3：共變異數矩陣與自由度皆相同($\Sigma_i = \Sigma$， $v_i = v$)。

爲了估計此模型參數的 MLE，先建立在此參數結構下的 EM 演算法。令待估參數 $\theta = (\mu_1, \ldots, \mu_g, \Sigma, v)$ ，在 EM 演算法中，一樣考慮完整資料 $Y_c = (Y, \tau)$ ，其中

$$Y = \left(Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, \ldots Y_{i1}, \ldots, Y_{in_i}, \ldots, Y_{g1}, \ldots, Y_{gn_g}\right),$$
$$\tau = \left(\tau_{11}, \ldots, \tau_{1n_1}, \tau_{21}, \ldots, \tau_{2n_2}, \ldots \tau_{i1}, \ldots, \tau_{in_i}, \ldots, \tau_{g1}, \ldots, \tau_{gn_g}\right).$$

可以得到給定完整資料下 θ 的對數概似函數爲

$$\ell_c(\theta \mid Y_c) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[-\frac{p}{2}\log(2\pi) + \frac{p}{2}\log(\tau_{ij}) - \frac{1}{2}\log|\Sigma| - \frac{\tau_{ij}}{2}(Y_{ij} - \mu_i)'\Sigma^{-1}(Y_{ij} - \mu_i)\right.$$
$$\left. - \log\Gamma(\frac{v}{2}) + \frac{v}{2}\log(\frac{v}{2}) + \frac{v}{2}\left[\log(\tau_{ij}) - \tau_{ij}\right] - \log(\tau_{ij})\right].$$

同樣的，在 *t* 分佈下的層級結構利用貝氏定理，在第 i 個多變量 *t* 分佈母體下， $\tau_{ij} \mid Y_{ij}$ 的後驗分佈爲

$$\tau_{ij} \mid Y_{ij} \sim Gamma\left(\frac{v + p}{2}, \frac{(Y_{ij} - \mu_i)'\Sigma^{-1}(Y_{ij} - \mu_i) + v}{2}\right),$$

其中 $i = 1, \ldots, g$ ， $j = 1, \ldots, n_i$ 。

EM 演算法整理如下：

E 步驟：給定 $\theta = \hat{\theta}^{(k)}$ ，省略掉跟 θ 無關的項，可以得到 Q-函數爲

$$Q(\theta\,|\,\hat{\theta}^{(k)}) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[-\frac{1}{2}\log|\Sigma| - \frac{1}{2}\hat{\tau}_{ij}^{(k)}(Y_{ij}-\mu_i)'\Sigma^{-1}(Y_{ij}-\mu_i)\right.$$
$$\left. -\log\Gamma(\frac{v}{2}) + \frac{v}{2}\log(\frac{v}{2}) + \frac{v}{2}\left(\hat{k}_{ij}^{(k)} - \hat{\tau}_{ij}^{(k)}\right)\right],$$

其中

$$\hat{\tau}_{ij}^{(k)} = \frac{\hat{v}^{(k)}+p}{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)})+\hat{v}^{(k)}},$$
$$\hat{k}_{ij}^{(k)} = DG\left(\frac{\hat{v}^{(k)}+p}{2}\right) - \log\left(\frac{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)})+\hat{v}^{(k)}}{2}\right),$$

$DG(\alpha) = d\log\Gamma(\alpha)/d\alpha$ 爲雙伽瑪函數。

M 步驟：

1.固定 $\Sigma = \hat{\Sigma}^{(k)}$ 、 $v = \hat{v}^{(k)}$ 之下，讓 Q-函數最大化，可得到：

$$\hat{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n_i} Y_{ij}\hat{\tau}_{ij}^{(k)}}{\sum_{j=1}^{n_i}\hat{\tau}_{ij}^{(k)}}. \tag{6}$$

2.固定 $\mu_i = \hat{\mu}_i^{(k+1)}$ 、 $v = \hat{v}^{(k)}$ 之下，讓 Q-函數最大化，可得到：

$$\hat{\Sigma}^{(k+1)} = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\hat{\tau}_{ij}^{(k)}(Y_{ij}-\hat{\mu}_i^{(k+1)})(Y_{ij}-\hat{\mu}_i^{(k+1)})'}{n}. \tag{7}$$

3.固定 $\mu_i = \hat{\mu}_i^{(k+1)}$ 、 $\Sigma = \hat{\Sigma}^{(k+1)}$ 之下，讓 Q-函數最大化，可得到：

$$\hat{v}^{(k+1)} = \arg\max_v\left\{\left[\frac{v}{2}\log(\frac{v}{2}) - \log\Gamma(\frac{v}{2})\right] + \frac{v}{2}\left[\frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\hat{k}_{ij}^{(k)} - \hat{\tau}_{ij}^{(k)})}{n}\right]\right\}. \tag{8}$$

在最佳分類法則定理中，利用 EM 演算法得到 MLE 的參數估計，計算機率密度函數後，可以得到在此參數結構($\Sigma_i = \Sigma$， $v_i = v$)的樣本最佳分類法則：

分配觀測值 x_0 到 P_k ，假使

$$\pi_k f_k(x_0\,|\,\hat{\mu}_k, \hat{\Sigma}_k, \hat{v}_k)c(i\,|\,k) > \pi_i f_i(x_0\,|\,\hat{\mu}_i, \hat{\Sigma}_i, \hat{v}_i)c(k\,|\,i) \text{ ，對所有 } i \neq k \text{ 。} \tag{9}$$

其中 $f_k(\cdot\,|\,\mu_k, \Sigma_k, v_k)$ 爲第 k 群母體多變量 t 分佈的機率密度函數。當分錯成本相同時，此時樣本的最佳分類法則變成爲：

分配觀測值 x_0 到 P_k ，假使

$$\pi_k f_k(x) > \pi_i f_i(x) \text{，對所有} i \neq k \text{。} \tag{10}$$

參數結構 4：共變異數矩陣相同($\Sigma_1 = \cdots = \Sigma_g = \Sigma$)與自由度不相同。

如同 3.1.2 節中爲了估計此模型參數的 MLE，先建立在此參數結構下的 EM 演算法。假設待估參數$\theta = (\mu_1,\ldots,\mu_g, \Sigma, \nu_1,\ldots,\nu_g)$，在 EM 演算法中，一樣考慮完整資料$Y_c = (Y, \tau)$，所以可以得到完整資料的對數概似函數爲

$$\ell_c(\theta \mid Y_c) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[-\frac{p}{2}\log(2\pi) + \frac{p}{2}\log(\tau_{ij}) - \frac{1}{2}\log|\Sigma| - \frac{\tau_{ij}}{2}(Y_{ij}-\mu_i)'\Sigma^{-1}(Y_{ij}-\mu_i)\right.$$
$$\left. - \log\Gamma(\frac{\nu_i}{2}) + \frac{\nu_i}{2}\log(\frac{\nu_i}{2}) + \frac{\nu_i}{2}\left[\log(\tau_{ij}) - \tau_{ij}\right] - \log(\tau_{ij})\right].$$

綜合上述，EM 演算法整理如下：

E 步驟：給定$\theta = \hat{\theta}^{(k)}$，在省略掉跟$\theta$無關的項，可以得到 Q-函數爲

$$Q(\theta \mid \hat{\theta}^{(k)}) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[-\frac{1}{2}\log|\Sigma| - \frac{1}{2}\hat{\tau}_{ij}^{(k)}(Y_{ij}-\mu_i)'\Sigma^{-1}(Y_{ij}-\mu_i)\right.$$
$$\left. - \log\Gamma(\frac{\nu_i}{2}) + \frac{\nu_i}{2}\log(\frac{\nu_i}{2}) + \frac{\nu_i}{2}\left(\hat{k}_{ij}^{(k)} - \hat{\tau}_{ij}^{(k)}\right)\right],$$

其中

$$\hat{\tau}_{ij}^{(k)} = \frac{\hat{\nu}_i^{(k)} + p}{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)}) + \hat{\nu}_i^{(k)}},$$

$$\hat{k}_{ij}^{(k)} = DG\left(\frac{\hat{\nu}_i^{(k)} + p}{2}\right) - \log\left(\frac{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)}) + \hat{\nu}_i^{(k)}}{2}\right),$$

M 步驟：

1.固定$\Sigma = \hat{\Sigma}^{(k)}$、$\nu = \hat{\nu}_i^{(k)}$之下，讓 Q-函數最大化，得到更新的$\hat{\mu}_i^{(k+1)}$，同(6)。

2.固定$\mu_i = \hat{\mu}_i^{(k+1)}$、$\nu = \hat{\nu}_i^{(k)}$之下，讓 Q-函數最大化，得到更新的$\hat{\Sigma}^{(k+1)}$，同(7)。

3.固定$\mu_i = \hat{\mu}_i^{(k+1)}$、$\Sigma = \hat{\Sigma}^{(k+1)}$之下，讓 Q-函數最大化，可得到：

$$\hat{\nu}_i^{(k+1)} = \arg\max_{\nu_i}\left\{\left[\frac{\nu_i}{2}\log(\frac{\nu_i}{2}) - \log\Gamma(\frac{\nu_i}{2})\right] + \frac{\nu_i}{2}\left[\frac{\sum_{j=1}^{n_i}(\hat{k}_{ij}^{(k)} - \hat{\tau}_{ij}^{(k)})}{n_i}\right]\right\}. \tag{11}$$

在最佳分類法則定理中，此參數結構爲共變異數矩陣相同($\Sigma_1 = \cdots = \Sigma_g = \Sigma$)與自由度不相同的樣本最佳分類法則，同(9)與(10)。

參數結構 5：共變異數矩陣不相同與自由度相同($v_1 = \cdots = v_g = v$)。

同 3.1 節，待估參數$\theta = (\mu_1,\ldots,\mu_g, \Sigma_1,\ldots,\Sigma_g, v)$，在 EM 演算法中，完整資料的對數概似函數爲

$$\ell_c(\theta|Y_c) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[-\frac{p}{2}\log(2\pi) + \frac{p}{2}\log(\tau_{ij}) - \frac{1}{2}\log|\Sigma_i| - \frac{\tau_{ij}}{2}(Y_{ij}-\mu_i)'\Sigma_i^{-1}(Y_{ij}-\mu_i)\right.$$
$$\left. - \log\Gamma(\frac{v}{2}) + \frac{v}{2}\log(\frac{v}{2}) + \frac{v}{2}\left[\log(\tau_{ij}) - \tau_{ij}\right] - \log(\tau_{ij})\right].$$

EM 演算法整理如下：

E 步驟：給定$\theta = \hat{\theta}^{(k)}$，在省略掉跟$\theta$無關的項，可以得到 Q-函數爲

$$Q(\theta|\hat{\theta}^{(k)}) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[-\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}\hat{\tau}_{ij}^{(k)}(Y_{ij}-\mu_i)'\Sigma_i^{-1}(Y_{ij}-\mu_i)\right.$$
$$\left. - \log\Gamma(\frac{v}{2}) + \frac{v}{2}\log(\frac{v}{2}) + \frac{v}{2}\left(\hat{k}_{ij}^{(k)} - \hat{\tau}_{ij}^{(k)}\right)\right],$$

其中

$$\hat{\tau}_{ij}^{(k)} = \frac{\hat{v}^{(k)} + p}{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}_i^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)}) + \hat{v}^{(k)}},$$
$$\hat{k}_{ij}^{(k)} = DG\left(\frac{\hat{v}^{(k)}+p}{2}\right) - \log\left(\frac{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}_i^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)}) + \hat{v}^{(k)}}{2}\right).$$

M 步驟：

1.固定$\Sigma_i = \hat{\Sigma}_i^{(k)}$、$v = \hat{v}^{(k)}$之下，讓 Q-函數最大化，得到更新的$\hat{\mu}_i^{(k+1)}$，同(6)。

2.固定$\mu_i = \hat{\mu}_i^{(k+1)}$、$v = \hat{v}^{(k)}$之下，讓 Q-函數最大化，可得到：

$$\hat{\Sigma}_i^{(k+1)} = \frac{\sum_{j=1}^{n_i}\hat{\tau}_{ij}^{(k)}(Y_{ij}-\hat{\mu}_i^{(k+1)})(Y_{ij}-\hat{\mu}_i^{(k+1)})'}{n_i}. \qquad (12)$$

3.固定$\mu_i = \hat{\mu}_i^{(k+1)}$、$\Sigma_i = \hat{\Sigma}_i^{(k+1)}$之下，讓 Q-函數最大化，得到更新的$\hat{v}^{(k+1)}$，同(8)。

此參數結構爲共變異數矩陣不相同與自由度相同($v_1 = \cdots = v_g = v$)的樣本最佳分類法則，同(9)與(10)。

參數結構 6：共變異數矩陣與自由度皆不相同。

同前，待估參數$\theta = (\mu_1,\ldots,\mu_g, \Sigma_1,\ldots,\Sigma_g, v_1,\ldots,v_g)$，完整資料之對數概似函數爲

$$\ell_c(\theta|Y_c)=\sum_{i=1}^{g}\sum_{j=1}^{n_i}\Bigg[-\frac{p}{2}\log(2\pi)+\frac{p}{2}\log(\tau_{ij})-\frac{1}{2}\log|\Sigma_i|-\frac{\tau_{ij}}{2}(Y_{ij}-\mu_i)'\Sigma_i^{-1}(Y_{ij}-\mu_i)$$
$$-\log\Gamma(\frac{v_i}{2})+\frac{v_i}{2}\log(\frac{v_i}{2})+\frac{v_i}{2}\left[\log(\tau_{ij})-\tau_{ij}\right]-\log(\tau_{ij})\Bigg].$$

EM 演算法整理如下：

E 步驟：給定 $\theta=\hat{\theta}^{(k)}$ ，計算

$$\hat{\tau}_{ij}^{(k)}=\frac{\hat{v}_i^{(k)}+p}{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}_i^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)})+\hat{v}_i^{(k)}},$$
$$\hat{k}_{ij}^{(k)}=DG\left(\frac{\hat{v}_i^{(k)}+p}{2}\right)-\log\left(\frac{(Y_{ij}-\hat{\mu}_i^{(k)})'\hat{\Sigma}_i^{(k)^{-1}}(Y_{ij}-\hat{\mu}_i^{(k)})+\hat{v}_i^{(k)}}{2}\right).$$

Q-函數爲

$$Q(\theta|\hat{\theta}^{(k)})=\sum_{i=1}^{g}\sum_{j=1}^{n_i}\Bigg[-\frac{1}{2}\log|\Sigma_i|-\frac{1}{2}\hat{\tau}_{ij}^{(k)}(Y_{ij}-\mu_i)'\Sigma_i^{-1}(Y_{ij}-\mu_i)$$
$$-\log\Gamma(\frac{v_i}{2})+\frac{v_i}{2}\log(\frac{v_i}{2})+\frac{v_i}{2}\left(\hat{k}_{ij}^{(k)}-\hat{\tau}_{ij}^{(k)}\right)\Bigg].$$

M 步驟：

1.固定 $\Sigma_i=\hat{\Sigma}_i^{(k)}$ 、 $v=\hat{v}_i^{(k)}$ 之下，讓 Q-函數最大化，得到更新的 $\hat{\mu}_i^{(k+1)}$ ，同(6)。

2.固定 $\mu_i=\hat{\mu}_i^{(k+1)}$ 、 $v=\hat{v}_i^{(k)}$ 之下，讓 Q-函數最大化，得到更新的 $\hat{\Sigma}_i^{(k+1)}$ ，同(12)。

3.固定 $\mu_i=\hat{\mu}_i^{(k+1)}$ 、 $\Sigma_i=\hat{\Sigma}_i^{(k+1)}$ 之下，讓 Q-函數最大化，得到更新的 $\hat{v}_i^{(k+1)}$ ，同(11)。

此參數結構最佳分類法則，同(9)與(10)。

4. 模擬與實例分析

4.1 模擬分析

我們考慮兩個狀況下的模擬資料：(1)資料母體爲來自多變量常態分佈伴隨著隨機的雜訊；(2)資料母體爲來自多變量 *t* 分佈。

此模擬分析的目的爲利用前一節所建構在多變量常態分佈與多變量 *t* 分佈兩個母體假設下的區別分析。在 6 種參數結構(1)共變異數矩陣相同(equal Σ_i ，ES)；(2)共變異數矩陣不相同(unequal Σ_i ，US)；(3)共變異數矩陣及自由度相同(equal Σ_i ；equal v_i ，ESEDF)；(4)共變異數矩陣不相同但自由度相同(unequal Σ_i ；equal v_i ，USEDF)；(5)共變異數矩陣相同但自由度不相同(equal Σ_i ，unequal v_i ；ESUDF)；(6)共變異數矩陣及自由度皆不相同(unequal Σ_i ，unequal v_i ；USUDF)下去探討產生的模擬資料其分類的表現。

爲了計算方便，在此模擬分析中假設事前機率與錯誤分類的損失成本皆爲相同。此外，對於分錯率的計算，在此使用 leave-one-out (LOO)的技術，見文獻 Lachenbruch and Mickey (1968)。LOO 是將訓練資料分爲 *n* 等分，每次保留一筆資料作爲分類的測試資料，其餘資料用來建構各參數結構下的區別模型，利用前一節所敘述的分類法則將保留下來的測試資料重新分類，重覆測試 *n* 次測試其分錯率。

4.1.1 模擬分析一

首先，我們考慮來自於兩個母體爲二維度多變量常態分佈下所模擬生成 100 筆的樣本資料，其中每個母體生成 50 個樣本，同時參數給定爲：

$$\mu_1 = (a, 0)^T,\ \ \mu_2 = (-a, 0)^T,\ \ \Sigma_1 = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},\ \ \Sigma_2 = \begin{bmatrix} 2 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}.$$

其中 *a* 給定爲 3 與 1。此外，額外再加上 50 筆雜訊的樣本點其來自於範圍從–10 到 10 的均勻分佈中隨機取值，最後模擬後的 150 筆資料於圖 2 中呈現。此圖中來自兩群的資料代表符號爲＋與○，雜訊資料代表符號爲●，而橢圓代表二維多變量常態下 95%的信賴區域。

在此模擬中 50 筆雜訊的樣本點各分成 25 筆隨機置入由兩個二維常態分佈下各生成的 50 個樣本中，如此資料依然爲兩個群體而每個群體各有 75 筆資料樣本但其中 25 筆資料爲雜訊。在此操作下，爲了提高可信度，重覆模擬 20 次，並且在各種參數結構狀況下做區別分析，執行 LOO 計算分錯率，再計算其平均，結果見表 1。

從表 1 中可以發現在加入雜訊的模擬資料以母體爲 *t* 分佈架構下的分錯率明顯小於在母體假設爲常態分佈，特別的是當 $a=1$時，即當此模擬資料中兩群的平均中心越接近時，在 *t* 分佈架構下區別分析的分錯率表現越好。同時也看出在常態分佈架構下的區別分析模型容易被雜訊或離群值所影響，相反的具有穩健性質的 *t* 分佈由於多了自由度參數的調整使其影響較小。

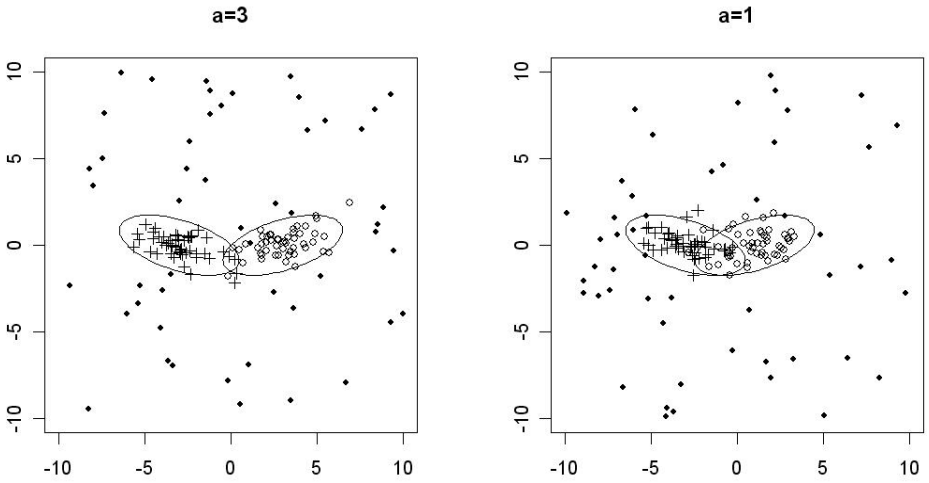

圖 2. 模擬資料散佈圖

表 1. 常態分佈與 *t* 分佈分錯率(%)之比較

參數值	MVN		MVT			
	ES	US	ESEDF	USEDF	ESUDF	USUDF
a=3	20	25.33	16.67	17.33	16.67	16.67
a=1	35.33	45.33	32	30.67	28	31.33

4.1.2 模擬分析二

我們模擬來自 *t* 分佈的資料去探討常態分佈與 *t* 分佈中在各種不同參數結構下區別分析中分錯率的表現。考慮有兩個母體，來自於爲二維多變量 *t* 分佈，其中參數 μ_i 與 Σ_i ($i=1,2$)的設定如前一次的模擬分析，而自由度設定爲 $v_1=v_2=2$。每個母體生成 100 個樣本，總共 200 個樣本並且重覆模擬 20 次產生此模擬的資料，在各種參數結構狀況下做區別分析，執行 LOO 計算其分錯率，再計算其平均，結果如表 2 所呈現。

表 2. 常態分佈與 t 分佈分錯率(%)之比較

參數值	MVN		MVT			
	ES	US	ESEDF	USEDF	ESUDF	USUDF
a=3	25.85	24.87	25.45	24.12	24.95	23.7
a=1	32.17	31.15	31.4	27.45	30.85	27.07

表 2 中可以顯然看出母體假設爲常態分佈的區別分析模型在此模擬中分錯率的表現並沒有相當顯著，意味著當資料不具有常態分佈的特性時，如某些極端的離群值會造成在區別分析上分類錯誤的結果。而 *t* 分佈在不同參數結構下的平均分錯率皆小於常態分佈；一樣地，當 $a=1$，即當 2 個母體所生成的資料平均中心越接近時，*t* 分佈下的平均分錯率相對於常態分佈的表現都有較大的差異。

4.2 實例分析

4.2.1 實例分析一

考慮牙齒資料，此資料最初由 Potthoff and Roy (1964)兩位學者研究分析，接著 Lee and Geisser (1975)，Fearn (1975)，Rao (1987)以及 Lee (1988, 1991)等學者考慮在成長曲線模型下對此資料分析與驗證模型的預測能力，並同時考慮利用統計檢定對成長曲線模型不同共變異數矩陣結構做選擇。

此資料中總共測量了 11 位女生與 16 位男生在 8 歲、10 歲、12 歲與 14 歲等不同的觀察時間點測量由腦垂體的中心到翼上頜裂橫的距離，單位爲公釐。假設事前機率與錯誤分類的損失成本皆爲相同，在母體爲多變量常態以及多變量 *t* 分佈假設下，探討學童性別於 6 種參數結構下分類的表現。

表 3 中可以顯然看出在多變量 *t* 分佈假設下區別分析分錯率的表現比多變量常態假設下有

很明顯的差異，特別的是當多變量 *t* 分佈共變異數矩陣與自由度不相同時分錯率最小，意味著牙齒資料在此參數結構下的區別能力是最好的。

表 3. 常態分佈與 t 分佈分錯率(%)與 BIC 之比較

常數	MVN		MVT			
	ES	US	ESEDF	USEDF	ESUDF	USUDF
分錯率	25.9	18.5	25.9	14.8	14.8	11.1
$\ell(\hat{\theta}\|Y)$	–208.41	–196.53	–198.43	–194.66	–195.88	–192.11
BIC	–476.15	–485.34	–459.49	–484.91	–457.68*	–483.10

$\ell(\hat{\theta}|Y)$ 表示對數概似函數之最大值；*表示 BIC 法則選取的最佳模式。

此外，同時也考慮用貝氏訊息準則(BIC; Schwarz, 1978)去驗證最合適的模型，BIC 公式計算如下：

$$\text{BIC} = 2\ell_{\max} - m\log n,$$

其中 $\ell_{\max}$ 表示最大的對數概似函數值， m 代表模型的參數個數， n 代表資料的樣本數。若 BIC 值愈大，表示該模型較合適。表 3 中可以知道當在多變量 *t* 分佈共變異數矩陣相同與自由度不相同時 BIC 值爲最大，代表在此資料下多變量 *t* 分佈的模型假設比多變量常態分佈的模型較適合。

4.2.2 實例分析二

爲了更進一步說明以多變量 *t* 分佈爲基礎的區別分析其穩健的性質，我們分析 Campbell and Mahon (1974)雜色瘦方蟹的資料。考慮 100 隻藍色螃蟹，其中有 50 隻爲公螃蟹與 50 隻爲母螃蟹。每個樣品共有 5 個測量變數(單位爲公釐)，分別爲：額唇的寬度(FL)、後背的寬度(RW)、甲殼中線長度(CL)、甲殼最大寬度(CW)與身體高度(BD)。資料散佈圖矩陣如圖 3，依性別此資料被分成兩群：公螃蟹代表符號爲○，標記紅色；母螃蟹代表符號爲△，標記藍色。此筆資料在 Peel and McLachlan (2000)與 Lin et al. (2004)已證明了在混合 *t* 模型下有較穩健的分群表現。

我們仿照 Peel and McLachlan (2000)在第 2 個變數的第 25 個觀察值加上特定的擾動值(± 5、± 10 、± 20)，一樣地，假設事前機率與錯誤分類的損失成本皆爲相同，在母體爲多變量常態以及多變量 *t* 分佈假設下與擾動值的情況下，探討藍色螃蟹性別於 6 種參數結構下分類的表現。

表 4 中顯然可以觀察到當加入的擾動值越大，在多變量常態分佈假設下的區別分析當參數結構共變異數矩陣相同時，其分錯率明顯偏高；相反的，在多變量 *t* 分佈假設下的區別分析不管加入的擾動值爲何，不同參數結構下的分錯率表現相對的變動並不大。

依照 BIC 的選模準則，在不同的擾動值下最大的 BIC 值總是發生在多變量 *t* 分佈的模型假設下，代表加入擾動值後的螃蟹資料在多變量 *t* 分佈的模型假設下較適合，也意味著多變量 *t* 分佈較不受資料離群值的影響，在估計及分群表現上較具穩健特性。

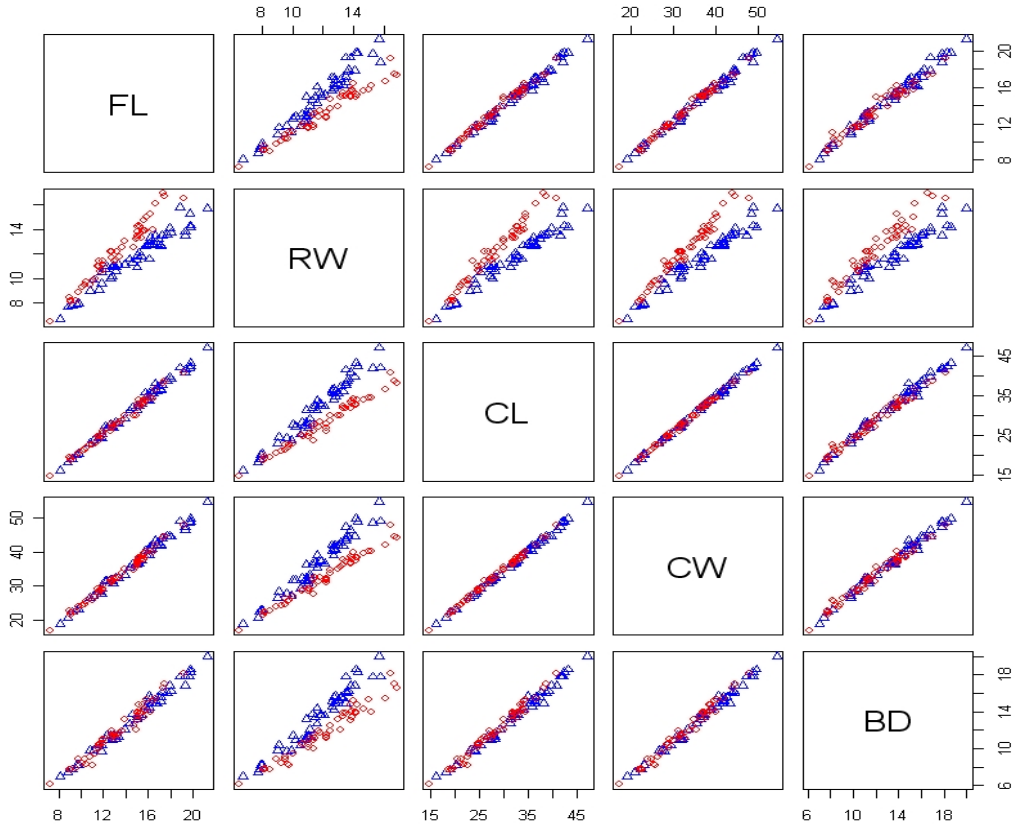

圖 3. 藍色螃蟹資料散佈圖

表 4. 多變量常態分佈多變量與 *t* 分佈分錯率(%)與 BIC 之比較

常數		MVN		MVT			
		ES	US	ESEDF	USEDF	ESUDF	USUDF
+20	分錯率	24	10	10	5	6	7
	BIC	–1410.20	–1311.87	–1221.63	–1203.07	–1217.47	–1197.62*
+10	分錯率	9	7	10	5	6	7
	BIC	–1291.81	–1246.05	–1206.01	–1187.41	–1204.14	–1184.63*
+5	分錯率	8	7	9	6	8	6
	BIC	–1207.65	–1187.97	–1188.39	–1169.80*	–1189.27	–1170.20
–5	分錯率	7	6	9	5	6	6
	BIC	–1205.58	–1185.05	–1187.99	–1168.61*	–1188.98	–1169.21
–10	分錯率	9	7	9	5	6	6
	BIC	–1290.03	–1244.24	–1205.85	–1186.89	–1204.03	–1184.19*
–20	分錯率	18	9	8	5	6	7
	BIC	–1409.12	–1310.91	–1221.56	–1202.83	–1217.42	–1197.42*

*表示 BIC 法則所選取的最佳模式。

5. 結論

本文中，在具穩健性質的多變量 *t* 分佈假設下去建構新的分類法則。在所提出的分類法則中，考慮不同的參數結構，並且以 EM 演算法去估算未知母體參數的最大概似估計值。

在模擬研究中，利用區別分析判別樣本資料的類別，藉由計算其分錯率大小來比較多變量常態分佈與多變量 *t* 分佈的表現。我們發現在多變量 *t* 分佈的假設下分錯率比多變量常態分佈低。在牙齒資料的實例分析中，多變量 *t* 分佈的分錯率的表現比多變量常態分佈有明顯差異。當參數結構爲共變異數矩陣與自由度皆不相同時分錯率最小。除此之外，在藍色螃蟹資料的實例分析中，也發現加入擾動值時多變量 *t* 分佈的分錯率較多變量常態分佈低。因此，當母體有異常值時，相較於多變量常態分佈，利用具穩健性質的多變量 *t* 分佈來執行統計分析可以獲得充分解釋資料的能力。在未來研究上，將發展兩群與多群多變量偏斜常態母體與多變量偏斜 *t* 母體之分類法則，提供研究者在日後的資料分析與建構分類法則上有更多的選擇。

致謝

本文作者誠摯感謝主編邀稿以及編輯委員與匿名審稿者所提供之寶貴意見，使文章之完整性與可讀性更臻完善。此外，本文研究成果係由國科會專題研究計畫 NSC 99-2118-M-005-001-MY2 提供部分之經費補助完成，特此致謝。

參考文獻

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, Third Edition, John Wiley, New York.

[2] Campbell, N. A. and Mahon, R. J. (1974). A multivariate study of variation in two species of rock crab of the genus Leptograpsus. *Australian Journal of Zoology*, 22, 417-455.

[3] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

[4] Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika*, 62, 89-100.

[5] Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, Sixth Edition, Prentice Hall, New Jersey.

[6] Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*, Third Edition, Cambridge University Press, New York.

[7] Lachenbruch, P. A. and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics*, 10, 1-11.

[8] Lee, J. C. and Geisser, S. (1975). Applications of growth curve prediction. *Sankhya, Series A*, 37, 239-256.

[9] Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association*, 83, 432-440.

[10] Lee, J. C. (1991). Test and model selection for the general growth curve model. *Biometrics*, 47, 147-159.

[11] Lin, T. I., Lee, J. C. and Ni, H. F. (2004). Bayesian analysis of mixture modeling using the multivariate t distribution. *Statistics and Computing*, 14, 119-130.

[12] Liu, C. (1997). ML estimation of the multivariate t distribution and the EM algorithms. *Journal of Multivariate Analysis*, 63, 296-312.

[13] Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

[14] Peel, D. and McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, 10, 339-348.

[15] Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science*, 2, 434-471.

[16] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.

[17] Sutradhar, B. C. (1990). Discrimination of observations into one of two t populations. *Biometrics*, 46, 827-835.